Finch Evolution Lab – 65 Points

A Model for AP Biology Statistical Practices

Data analysis is a major part of all science practices. Choosing which method to use for analyzing data is an important skill to practice, and this lab will further your understanding of the different statistical methods.

A. Using Mean, Median, Mode & Range

- Mean – Useful for comparing averages of data sets when data is distributed normally.
- Median – Useful for comparing averages of data sets when data is ***not*** distributed normally (outliers skew the mean)
- Range – Useful for understanding spread of the data.
- Mode – Useful for understanding any pattern of repetition.

B. Populations versus Samples

Knowing the true population size and being able to use all members in an experiment is usually impossible. We instead take **samples** from a population with the understanding that our samples will lead to **some error** since they may not be representative of the **ACTUAL** population.

C. Hypothesis Testing

In this course the experiments are designed to test **null hypotheses (H_0)**. These state "There will be no difference in ___ and ___" or can be stated that the outcomes should be the same. Once a null hypothesis is stated, it can be statistically tested for significance using the experimental data. At the AP & College level, a visual interpretation of data will not suffice. If a null hypothesis is rejected, you accept the **alternate hypothesis (H_a)**, stating there is a statistically significant difference between your data sets not due to chance/error.

D. Statistical Tests

There are MANY statistical tests, but you must learn & utilize 2 in this course:

1. **The t-Test** – Tests for a statistically significant difference between 2 data sets or one data set. This test is mostly used when there are numerous data for each group, being tested as different from each other. The results of these tests use standard error & standard deviation, which will be discussed in a later section.
    a. After performing the t-Test, a **p-value** will be generated. This value represents the likelihood that your data are different due to error/chance. **If the value is less than .01, we reject the null hypothesis that they are the same and accept the alternate hypothesis meaning they are different statistically.**

2. **The Chi Square Test** – Tests for a statistically significant difference between outcomes of data that we **observe** by an experiment and those we calculate **expected** values for based on the hypotheses. This test is mostly used when there are at least 2 outcomes for data to be organized into.
    a. After performing the Chi Square test, the value known as X^2 will be the number of interest. If this value is greater than that of the appropriate corresponding value in the Chi Square table, **we reject the null hypothesis that they are the same and accept the alternate hypothesis meaning they are different statistically.**

E. Graphs

You are expected to use the correct type of graph for a given data set. Some guidelines are below:

1. Line Graphs: Used to express trends over time or another variable that changes as the other variable changes.
2. Bar Graphs: Used to express categorical data into sets; useful for seeing how multiple classes of data compare to one another.
3. Pie Graphs: Used to compare parts to a whole.

F. Types of Error Representation

As discussed in the populations vs. samples section, there will always be error in data collection due to the sample not being exactly the same as the population. Two ways of reporting error are by the Standard Deviation & the Standard Error.

1. Standard Deviation is used in error reporting when data being analyzed are focused on **how much variance there is in the data**. The higher the value, the more variance there is. A standard deviation is calculated for each data set and these are used to determine the likelihood that 2 data sets are different due to chance.
2. Standard Error is used in error reporting when data being analyzed are focused **on how much the means of different trials of a single experiment vary from each other**. Basically this shows whether or not your samples are a reliable representation of the true population mean.

Notice that both of these error representations involve the mean. If the results of a statistical test don't match up to what you might think they should be, these are good ways to indicate perhaps the mean is not a good way to analyze data and instead the median should be used.

In any case, error bars can be generated on graphs using a computer application or by hand and are able to be reported as standard deviation or standard error. A bar is drawn above the mean/median by 1x the standard error or standard deviation; repeat for below the mean/median.

The following pages contain case study data which you will practice using statistical analysis with.

Case Study: The Evolution of Finches

A. The warbler finch eats seeds for the main part of its diet. A scientist wanted to research how the seed characteristics influenced the finch's evolution. Finches with thicker beaks are more able to crack open harder seeds than those with thinner beaks. High precipitation tends to make seeds develop softer than areas with low precipitation which produce hard seeds. 2 different samples of finches were captured from each population location, A & B, their beaks were measured, and then they were released back to their appropriate location. The data tables below represent the data collected by the scientist:

Table 1: Population A-1 sample data of beak thickness in mm

3.15	5.14	5.24
5.09	5.15	5.29
5.11	5.23	5.31
5.11	5.23	5.32
5.12	5.24	5.32

SD _____

SE _____

Table 2: Population A-2 sample data of beak thickness in mm

4.99	5.10	5.21
4.99	5.12	5.26
5.01	5.14	5.30
5.06	5.18	5.36
5.09	5.19	5.37

SD _____

SE _____

Table 3: Population B-1 sample data of beak thickness in mm

7.02	7.13	7.23
7.11	7.13	7.23
7.11	7.13	7.24
7.12	7.15	7.24
7.12	7.16	7.26

SD _____

SE _____

Table 4: Population B-2 sample data of beak thickness in mm

7.24	7.25	7.29
7.24	7.26	7.29
7.24	7.28	7.30
7.25	7.28	7.31
7.25	7.28	7.31

SD _____

SE _____

Calculate the standard deviation (SD) & standard error (SE) for each data set above (8).

1. **Justify** which sample would be best represented using the median instead of the mean. (2)

2. For each question below, **state** an appropriate null hypothesis & alternate hypothesis, **explain** which statistical test is the most appropriate and which error reporting method is best.

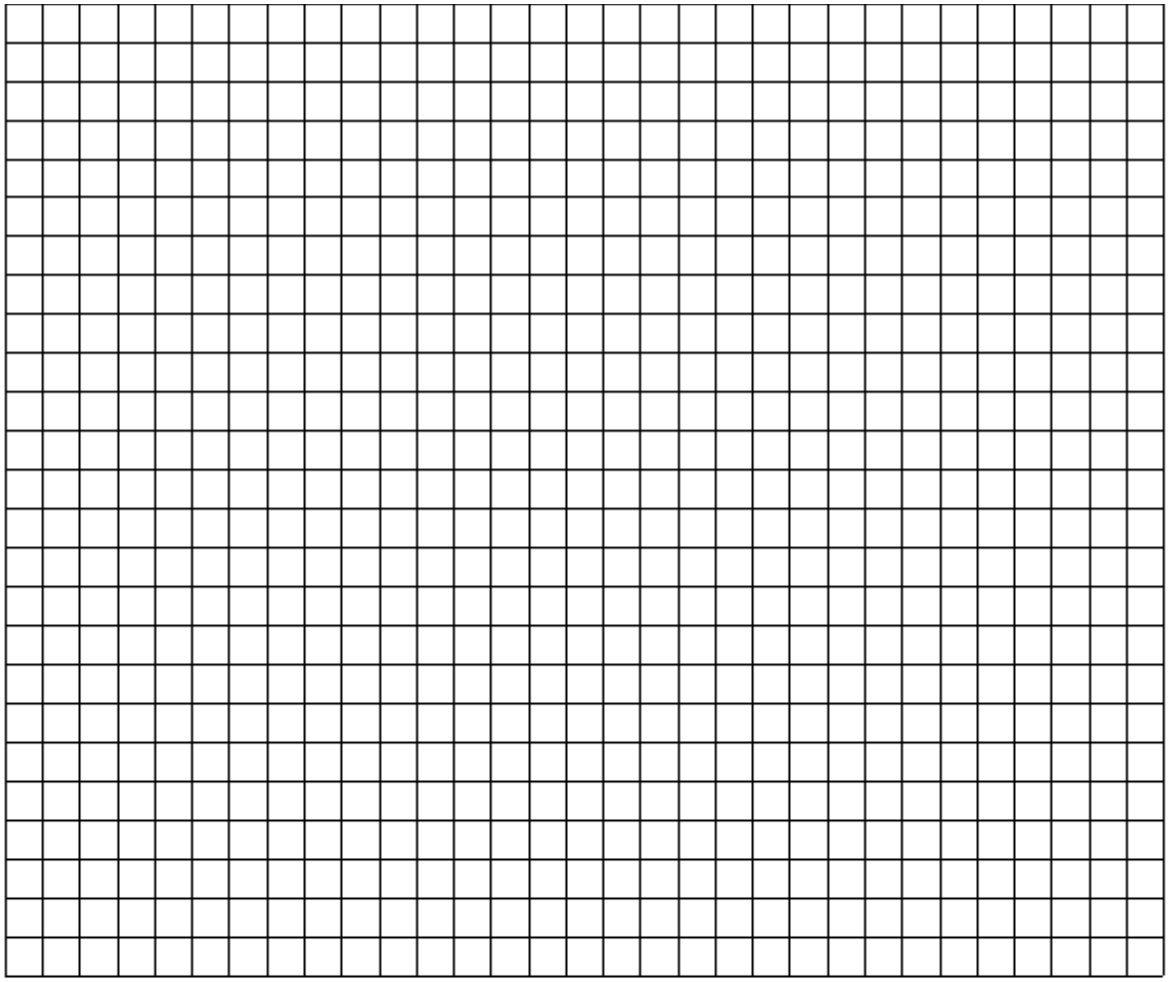
a. Are the finch beaks different between the 2 samples of Population A? (6)

b. Is the population location a significant factor in determining beak size? (6)

c. Is there an even distribution into 3 intervals of finch beaks in population A-2? (6)

3. Construct an appropriate graph below to answer the question in 2b based on the **first** sample sets.

Be sure to label all relevant parts of a graph & error analysis. (10)



Record the statistical test value _____

Explain whether to accept or reject the null hypothesis. (2)

4. Perform the appropriate statistical test for 2c (10)

State whether to accept or reject the null hypothesis based on the test statistic (2)

5. Based on all the data analyzed, describe the best **conclusions (data findings)** for each hypothesis in #2 (4)

6. Based on all the data analyzed, describe the best **biological significance (relevance to scientific community)** of the finch study. Your response should discuss the finch's evolution as related to natural selection. (5)

7. Based on all the data analyzed, describe the best **refinements** for this experiment in future studies. (4)